

# 병원 ERP시스템을 적용한 CEP 기반 실시간 분석시스템 개발

김미진 · 유윤식 · 서영우\* · 장중욱\*\*

동의대학교 부산IT융합부품연구소 · \*㈜투그램시스템즈 · \*\*동의대학교

## Development of CEP-based Real Time Analysis System Using Hospital ERP System

Mi-Jin Kim · Yun-Sik Yu · Young-Woo Seo\* · Jong-Wook Jang\*\*

Convergence of IT Devices Institute Busan, Dong-Eui University · \*Toogram Systems Inc. · \*\*Dong-Eui University

E-mail : agicap@deu.ac.kr, ysyu@deu.ac.kr, youngwoo@toogram.com\*, jwjang@deu.ac.kr\*\*

### 요 약

개개인의 데이터가 비즈니스적으로 중요하지 않을 수 있지만, 대량으로 모으면 그 안에 숨겨진 새로운 정보를 발견할 가능성이 있는 데이터의 집합체로 빅데이터 분석 활용 사례는 점차 늘어나는 추세이다. 빅데이터 분석 기술 중 전통적인 데이터 분석방법인 하둡(Hadoop)은 예전부터 현재에 이르기까지 정형·비정형 빅데이터 분석에 널리 사용되고 있는 기술이다. 하지만 하둡은 배치성 처리 시스템으로 데이터가 많아질수록 응답 지연이 발생할 가능성이 높아, 현재 기업 경영환경과 시장환경에 대한 엄청난 양의 고속 이벤트 데이터에 대한 실시간 분석이 어려운 상황이다.

본 논문에서는 급변하는 비즈니스 환경에 대한 대안으로 오픈소스 CEP(Complex Event Processing) 기반 기술을 사용하여 초당 수백에서 수십만건 이상의 이벤트 스트림을 실시간으로 지연 없이 분석 가능하게 하는 실시간 분석 시스템을 개발하여 병원 ERP시스템에 적용하였다.

### 키워드

빅 데이터, CEP, 실시간 분석시스템, 병원, ERP

### 1. 서 론

스마트 시대의 소셜 네트워크, 사물인터넷, 라이프로그 데이터 등은 빅데이터 시대 진입에 중요한 요소들이다. 스마트 단말기는 수많은 데이터를 생산하고, 그 기기들로부터 생산되는 수많은 데이터들은 분산 파일 형태로 수집되어 중요한 정보로 가공된다. 빅데이터는 지금까지 이해할 수 없었던 정보를 이해하고, 분석할 수 없었던 대용량의 비정형 데이터를 처리하는 기술로서 데이터를 이용한 지능형 서비스 구현의 기반 기술이다. 이 기술은 가까운 미래에 사용자에게 제공하는 서비스의 질을 높이고 전문가의 역할과 가치를 바꿔놓을 것이다[1]. 구글, 페이스북, 아마존, 야후 등의 소셜 미디어나 인터넷 업체들이 빅데이터 기반의 소셜 분석 효과를 입증하면서 국내외에서 빅데이터를 활용한 교통시스템, 수자원시스템, 방법시스템, 의료시스템 등의 공공서비스 운영을 도모하고 있으며, 민간기업에서도 기업경영개선과

마케팅 효율화를 도모하여 수익성 향상과 프로세스 효율화는 물론 빅데이터 기반의 새로운 비즈니스 모델 개발을 서두르고 있다.

개개인의 데이터가 비즈니스적으로 중요하지 않을 수 있지만, 대량으로 모으면 그 안에 숨겨진 새로운 정보를 발견할 가능성이 있는 데이터의 집합체로 빅데이터 분석 활용 사례는 점차 늘어나는 추세이다. 빅데이터 분석 기술 중 전통적인 데이터 분석방법인 하둡은 예전부터 현재에 이르기까지 정형·비정형 빅데이터 분석에 널리 사용되고 있는 기술이다. 하지만 하둡은 배치성 처리 시스템으로 데이터가 많아질수록 응답 지연이 발생할 가능성이 높아, 현재 기업 경영환경과 시장환경에 대한 엄청난 양의 고속 이벤트 데이터에 대한 실시간 분석이 어려운 상황이다. 또한, 다양한 기업군에서 빅데이터 분석 정보를 활용하길 원하지만, 빅데이터 솔루션은 대부분 고비용의 문제로 도입해서 사용하기 힘든 실정이다.

본 논문에서는 위에서 언급한 부분들을 해결할

수 있는 대안으로 오픈소스 CEP 기반 기술을 사용하여 저비용으로 엄청난 양의 고속 이벤트 스트림을 실시간으로 지연 없이 분석 가능하게 하는 실시간 분석시스템을 구축하였다. 또한, 다양한 기업군 중에서 중·소규모의 병원을 대상으로 병원 ERP 시스템에 적용하여 환자 중심의 정보와 의료장비 데이터를 분석함으로써 효율적인 환자 관리 및 병원의 경영관리를 체계적으로 제공할 수 있는 시스템을 구현하였다.

## II. 관련연구

### 2.1 빅데이터(Big Data)

빅데이터란 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술[2]을 의미한다.

이러한 빅데이터의 속성은 크게 3V로 언급하며 크기(Volume), 속도(Velocity), 다양성(Variety)을 이른다. 최근에는 가치(Value)나 복잡성(Complexity)을 덧붙이기도 한다. 그러나 빅데이터의 활용적 측면에서 바라보았을 때 중요한 속성 중 하나가 바로 ‘저비용(Low Cost)’이다. 빅데이터의 태생적 근원이 기존의 그것과는 다른 체계에서 비교적 저렴한 비용으로 저장되고 처리되길 바라는 것이었기 때문이다[3].

빅데이터의 기술은 크게 데이터 수집기술, 저장기술, 처리기술, 분석기술, 표현 및 활용기술, 관리 (Infra, Biz)기술 분야로 나눌 수 있다. 각 기술 분야에서 요구되는 기술 및 문제점들을 보면, 수집 기술에서는 대용량의 데이터의 적재시간이 전체시간의 대부분을 차지하며 데이터가 지속적으로 증가한다는 점이고, 저장기술에서는 데이터 저장과 관리에 고비용이 든다는 점이다. 또한 처리/분석 기술 등에서도 처리 및 연산에 높은 비용이 소요되며, 장시간의 처리로 인해 데이터의 적시성이 부족하다는 점 등의 문제점들이 거론되고 있다.

### 2.2 하둡(Hadoop)

하둡은 현재 빅데이터 처리에 가장 선호되고 있는 전통적인 분산처리기술을 핵심으로 하는 솔루션이다. 비교적 단순한 프로그래밍 모델을 사용하여 대용량의 데이터를 분산처리하기 위한 Apache 오픈소스 프로젝트로 자바 기반의 프레임워크다[4]. 이것은 야후와 페이스북 등에서 핵심 기술로 사용되고 있으며, 많은 기업들이 자사 솔루션에 응용하여 사용하고 있다. 주요 구성은 하둡 분산 파일 시스템인 HDFS(Hadoop Distributed File System)와 MapReduce라는 분산처리 시스템으로 구성된다. HDFS는 GFS(Google File System)를 모델로 해서 만들어진 오픈 소스이다. 따라서 HDFS는 GFS와 동일한 특성을 가진다. HDFS는 대용량의 파일을

청크(chunk, 64MB)라는 단위로 분할하여 데이터 노드에 3개씩 분산 저장된다. 청크가 어느 데이터 노드에 저장되었는지에 대한 메타데이터는 네임노드에 저장된다. 구동 방식은 HDFS나 MapReduce는 Namenode와 Datanode라 불리는 마스터와 다중 슬레이브 구조로 이루어져 있다. 그림 1의 Namenode는 파일시스템의 Metadata를 관리하고 Client와 Datanode 사이의 데이터 I/O를 제어한다. Datanode는 실제 데이터를 저장하고 있으며 Client와의 데이터 I/O 및 블록 복사를 직접 수행한다.

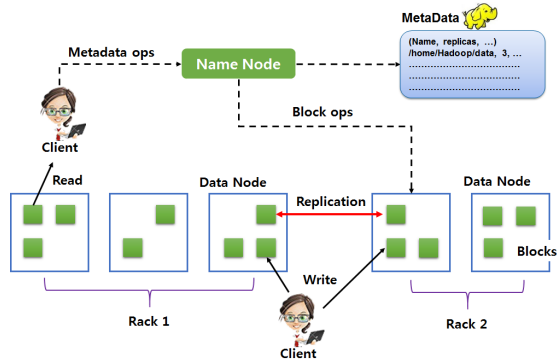


그림 1. HDFS 구조

MapReduce 프레임워크는 데이터 처리 시간이 오래 걸리던 대용량 데이터를 단시간 내에 일괄 처리로 연산시킬 수 있는 분산/병렬처리용 프로그래밍 모델이다. 대용량 데이터로부터 Key-Value에 대한 커스터마이징을 지원하며, Binary 탐색, Hash 알고리즘 적용 등을 기반으로 데이터를 처리를 고속으로 실행시켜준다. 이러한 특성에 따라 하둡은 로그 데이터와 대용량 멀티미디어 데이터의 저장 및 분석 등의 분야에 활용되고 있다[5].

### 2.3 CEP(Complex Event Processing)

CEP는 복합 이벤트 처리 기술로 여러 이벤트 소스로부터 발생한 이벤트를 대상으로 실시간으로 의미 있는 데이터를 추출하여 대응되는 액션을 수행하는 것을 말한다[6]. 이때 이벤트 데이터는 스트림 데이터로서 대량으로 지속적으로 입력되는 데이터, 시간 순서가 중요한 데이터, 끝이 없는 데이터이다. 이러한 스트림 데이터는 전통적인 관계형 데이터베이스에서는 실시간 처리 및 분석이 불가능하다. CEP는 바로 이런 스트림 데이터를 실시간으로 분석하는 이벤트 데이터 처리 솔루션이다. 즉, 데이터베이스나 파일, 하둡에 저장하지 않고, 다양한 고속의 이벤트 스트림을 In-Memory 기반으로 초당 수백/수십만건의 실시간 처리가 가능하다.

CEP 처리를 위해서는 CEP 엔진을 사용한다. CEP 엔진은 여러 시스템에서 발생하는 이벤트를 지정하고, 추출하고자 하는 이벤트 패턴을 등록하면, 여러 가지 이벤트 스트림들을 정제(filtering), 합산(aggregation), 수집(gather), 결합(join)한 후

패턴 매칭을 통하여 원하는 이벤트 패턴의 발생을 감지하는 기능을 제공한다.

CEP 기반의 분석방법은 기본적으로 Cisibility → Understanding → Get Insight 형태의 분석방법으로 기업 내 발생하는 실시간 데이터에 대한 가시화를 통해 해당 이벤트들간의 연관성 및 패턴을 이해하고 그 이해를 바탕으로 실시간 기업환경에 필요한 직관성을 얻음으로써 다양한 실시간 비즈니스에 대한 대응이 가능하다[7].

### 2.4 빅데이터 접근방식에 대한 하둡과 CEP

최근 빅데이터 기반의 분석 방법론은 저장 관점에서의 하둡 기반의 접근이 주된 접근 방법론이었던 것은 사실이다. 그러나 최근 그림 2처럼 빅데이터에 대한 접근은 하둡 생태계 기반의 배치분석 관점도 물론 중요한 부분이지만, 가트너 보고 자료처럼 실제적으로 CEP 기반의 실시간 분석 관점의 접근도 나날이 중요성을 더해가고 있다.

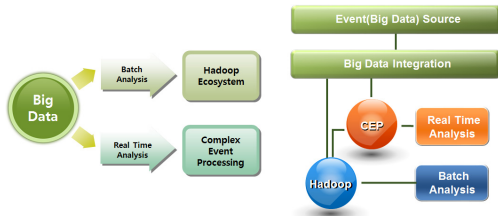


그림 2. 빅데이터 분석 관점에 대한 하둡과 CEP

기존의 DB, 하둡 기반의 전통적인 배치 처리와 다양한 고속의 이벤트 스트림을 In-Memory 기반의 Real-time 처리와의 분석 매커니즘의 차이점은 그림 3과 같다[8]. 본 논문에서는 선 분석, 후 저장 방식의 In-Memory 기반의 분석 처리 매커니즘을 사용하는 CEP 기술을 사용하였다.

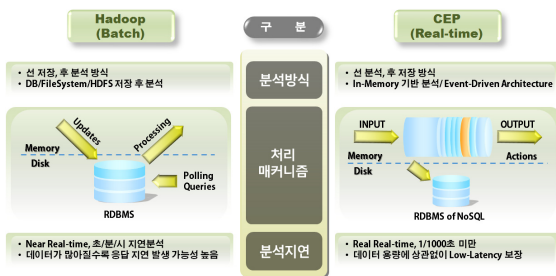


그림 3. 하둡과 CEP 분석 처리 매커니즘

### III. 시스템 구성 및 설계

CEP 기술은 적시 분석을 토대로 한 실시간 비즈니스 인텔리전스를 구현하는 가장 강력하고 단순한 방법으로 비즈니스 분야에서 부상하고 있는

며, 다양한 이벤트 처리 및 분석을 통해 실시간 모니터링, 조기 경보, 생산 현장 관리 등과 같은 새로운 가치를 제공한다.

본 논문에서는 빅데이터의 활용적 측면인 저비용을 고려하고 CEP 기반의 장점을 활용하여 아직 사례가 많지 않은 병원 ERP 시스템과 결합하여 빅데이터를 적용할 수 있는 실시간 분석 시스템을 구축하여 환자 중심의 정보와 병원장비 데이터 기반의 정보들을 가공·분석함으로써 효율적인 환자관리 및 의료기관의 경영관리를 위해 체계적이고 조직적인 환경을 제공하고자 하였다. 또한 주요 Adaptor 및 데이터 퍼블리셔/커스터마이징 기능을 구현하여 각 의료기관들의 요구에 따라 UI 화면 정의 및 개발이 가능하도록 하였다.

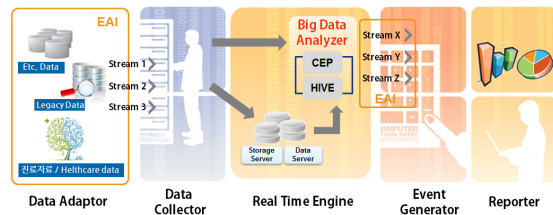


그림 4. CEP 기반 실시간 분석 시스템 아키텍처

실시간 분석 시스템 아키텍처는 그림 4와 같은 구성을 가진다. Data Adaptor를 통해 실시간 분석 시스템으로 유입·유출되는 여러 종류의 데이터들은 데이터의 protocol과 type 등 내·외부 이벤트 type으로 변환하여 전달하고, Data Collector에서 실시간 분석에 사용하는 스트림과 매핑 시키며 각 Event에서 유입되는 시스템을 NoSQL에 저장하는 기능도 함께 제공되어 시계열 분석 활용도 가능하다. NoSQL DB는 Cassandra를 통한 배치 이벤트처리를 하였다. Big Data Analyzer에서는 Data Collector에서 수집된 데이터를 오픈소스 기반의 실시간 분석 엔진 CEP와 Hive 기반의 배치레이어를 이용하여 데이터를 분석하고 Reporting 가능한 형태로 매핑하는 기능을 수행하고, Event Generator에서는 실시간 분석 결과를 사용자가 원하는 type과 protocol로 변환하여 제공하는 기능과 분석결과를 실시간으로 데이터베이스에 저장하는 기능도 수행한다. Reporter에서는 실시간 분석 결과에 대한 Dashboard, Alarm, Analysis에 대한 스케줄링 지원 등 실 사용자가 CEP 기반 실시간 분석 시스템을 시각적으로 접근할 수 있는 Web 기반 분석 기능을 제공한다.

### IV. 시스템 구현

전체 시스템의 프로세스 처리 과정을 간략하게 설명하자면 그림 5에서와 같이 Legacy는 자체 업무 프로세스 중 발생한 로그사항을 저장하고, 처리 프로세스 중 주요 업무에 대해서는 실시간 분석 처리를 수행한다.

배치 레이어는 데이터 aggregation을 통해 데이터를 수집하고 분석 작업을 통해서 데이터를 추출한다. 추출된 결과에 따라 중간 저장소 보관 및 실시간 분석 요청이 이루어질 수 있다. 실시간 처리 레이어는 Legacy의 요청과 배치 레이어의 요청을 처리하며, 처리하는 과정 중 필요한 데이터를 중간저장소에 요청 및 결과를 저장한다. 처리 결과는 모니터링 시스템 API 호출이 이루어진다. Dashboard는 분석결과를 사용자에게 요청에 따라 데이터를 시각화 하며, Dashboard에서 처리하는 데이터는 중간 저장소에 보관된다.

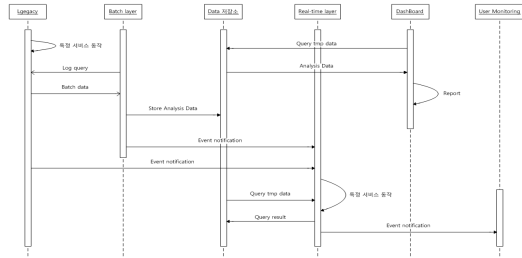


그림 5. 전체 시스템 프로세스 정의

그림 6은 설정된 Output Adaptor의 전체 데이터를 Dashboard 상에서 분석되는 모니터링 화면이다. 또 필터 기능을 통해 Server/Adaptor/Event/Hour 등을 설정하여 필요한 부분만 분석도 가능하다. 그림 7은 필터 기능을 적용한 모니터링 화면으로, 필터 적용은 Server는 210.109.9.114를, Output Adaptor는 Event를, Event는 event\_5를, Hour는 13:00 ~ 18:00를 선택했을 때의 분석 화면이다.

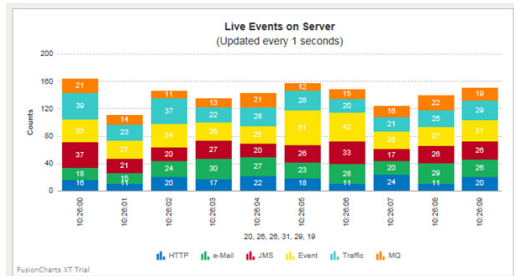


그림 6. 유입되는 전체 데이터 분석 모니터링 화면

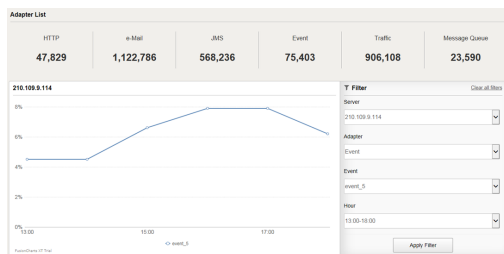


그림 7. 선택한 필터의 결과 화면

## V. 결론

맥킨지 보고서에 따르면 의료분야 가치는 국민 의료비의 절감과 혁신적인 임상연구를 가능케 하는 것과 연관이 깊은 것으로 조사 되었다. 이에 본 연구에서 병원 ERP 시스템을 적용한 CEP 기반 실시간 분석시스템 개발을 통해 급변하는 환경속에서 엄청난 양의 고속 이벤트 데이터 들을 실시간으로 처리할 수 있는 기반을 마련하였고, 보건 의료 분야에서의 빅데이터 분석의 폭넓은 활용에 적게나마 도움이 됨에 따라 사회경제적으로 큰 영향력을 발휘할 수 있을 것으로 생각된다.

개발된 빅데이터 분석시스템은 저비용으로 중소 규모의 병원에서도 활용될 수 있게 하였으며, 각 병원의 특정 요구에 더 빠르고 정확한 정보를 제공할 수 있고 효율적인 환자관리 및 경영관리를 통한 경제 가치 창출이 가능하다.

향후 병원의 데이터 활용에 대한 개발이 더욱 필요하겠고, 선박·물류 등 많은 기업에서 사용할 수 있는 추가적인 시스템 보완 등 지속적인 연구가 필요하다.

## ACKNOWLEDGMENT

본 논문은 정보통신산업진흥원에서 지원하고 (재) 부산정보산업진흥원이 주관하는 2014-15년 누리마루 R&BD 프로젝트 사업(SW융합 R&BD 지원 사업) 수행으로 연구하였음.

## 참고문헌

- [1] “2013년 IT산업 7대 Mega Trend” 한국정보산업연합회. 2013, 1장 ‘빅데이터 정책’
- [2] John Gantz & David Reinsel, << Extracting Value from Chaos >>, IDC IVIEW June, (2011년), p.6.
- [3] Philip Russom, “Big Data Analytics”, TDWI Research Fourth Quarter, p.6., 2011
- [4] K.-W. Park, K.-J. Ban, S.-H. Song, and E.-K. Kim, “Cloud-based Intelligent Management System for Photovoltaic Power Plants,” J. of the Korea Institute of Electronic Communication Sciences, vol. 7, no. 3, June 2012, pp. 591-596.
- [5] [http://hadoop.apache.org/docs/hdfs/current/hdfs\\_design.html](http://hadoop.apache.org/docs/hdfs/current/hdfs_design.html)
- [6] Luckham, David C., and Brian Frasca. Complex event processing in distributed systems. Computer Systems Laboratory Technical Report CSL-TR-98-754. Stanford University, Stanford 28, 1998.
- [7] <http://blog.naver.com/tibcokorea/20207252684>
- [8] [http://www.dbguide.net/knowledge.db?cmd=specialist\\_view&boardUid=180895&boardConfigUid=108&boardStep=0&categoryUid](http://www.dbguide.net/knowledge.db?cmd=specialist_view&boardUid=180895&boardConfigUid=108&boardStep=0&categoryUid), 이호철, No. 249, 2014. 7. 29